

# PLACEMENT OF SOUND SOURCES IN THE STEREO FIELD USING MEASURED ROOM IMPULSE RESPONSES<sup>1</sup>

*William D. Haines*

*Jesse R. Vernon*

*Roger B. Dannenberg*

*Peter F. Driessen*

Carnegie Mellon University, School of Computer Science  
Pittsburgh, PA, USA

University of Victoria  
Victoria, BC, Canada

{wdh, jvernon}@andrew.cmu.edu

rbd@cs.cmu.edu

peter@ece.uvic.ca

## ABSTRACT

Current advances in techniques have made it possible to simulate reverberation effects in real world performance spaces by convolving dry instrument signals with physically measured impulse response data. Such reverberation effects have recently become commonplace; however, current techniques apply a single effect to an entire ensemble, and then separate individual instruments in the stereo field via panning. By measuring impulse response data from each instrument's desired location, it is possible to place instruments in the stereo field using their unique initial reflection and reverberation patterns. A pilot study compares the perceived quality of dry signals convolved to stereo center, convolved to stereo center and panned to desired placement, and convolved with measured impulse responses to simulate actual placement. The results of a single blind study show a conclusive preference for location-based reverberation effects.

## 1. INTRODUCTION

When an ensemble performs on stage before a live audience, the audience's listening experience is theoretically enhanced by the stereo separation of the instruments as determined by their physical placement on stage. This effect does not occur by chance, as percussive instruments are often placed in the center of the stage, with bass and melodic instruments often separated to either side. The placement is formulated so as to reduce the effect of one instrument dominating the sound of another. Currently, when recording and mixing down albums, a single reverb is placed on each track, based upon either IIR filters or a convolution with a single measured impulse response. Placement is achieved using a combination of stereoscopic panning, pre-delays, decay times, and saturation levels in order to separate the

individual instrument tracks. This method is effective, but purely artificial, providing no real psycho-acoustical clues that the instrument field is properly placed.

When an instrument is played at one location on a stage versus another, the reverberation signature is different. This effect occurs because as sound radiates from the instrument, the sound energy reflects off of various walls, the floor, and ceiling, reaching one's ear at different time intervals and at different frequency dependent amplitudes. The effect is subtle, but, in principle, recognizable. Consequently, there is a unique impulse response associated with each location on the stage (paired with each listening location in the room). Theoretically, then, if each instrument signal in an ensemble is convolved with its unique location-based impulse response, then it should enhance the psycho-acoustical illusion of the separation of the instrument field, eliminating the need for artificial separation while still removing the perception of one instrument overpowering the others.

However, even convolution with impulse responses is only a simple approximation of sound radiation in a room. Acoustic instruments have frequency dependent radiation patterns that we do not model. The impulse responses used here incorporate the directional radiation patterns of the speakers used in the impulse response measurement process. These patterns will be different from those of acoustic instruments. Another limitation is that stereo recording does not capture the complex sound field available to the listener in an acoustic space. This is a fundamental limitation of the stereo format. Our goal is only to better simulate this format, not to overcome its limitations.

The extent to which the technique of virtual instrument placement via measured room impulse responses will improve the actual perceived quality of the performance is unknown; hence, the need for an appropriate study to

---

<sup>1</sup> Originally published as: William D. Haines, Jesse R. Vernon, Roger B. Dannenberg, and Peter F. Driessen, "Placement of Sound Sources in the Stereo Field Using Measured Room Impulse Responses," in *Proceedings of the 2007 International Computer Music Conference, Volume I*. San Francisco: The International Computer Music Association, (August 2007), pp. I-496 - 499.

evaluate the qualitative difference between current methods and the proposed method.

## 2. PREVIOUS RESEARCH

Current recording techniques are the culmination of many years of research and reasoning. Numerous studies have been conducted to evaluate the utility of current techniques, in addition to considering their ability to withstand the rigors of commercial practice. Formulations of the theory can be found in Pulkki among others [5]. Regarding virtual instrument placement via location-based reverberation, not much has been studied regarding the actual quality of the effect versus current methods. The theory behind the method has been outlined on several occasions, including discussions by Reller and Griesinger [6, 4]. The Roland SRV-330 Dimensional Space Reverb uses 24 early reflections to create the impression of a 3-D acoustic space [7]. However, actual quality perception tests and implementation detail are not available.

## 3. METHODOLOGY

### 3.1. Experimental Design

Given the timeframe of the study and our relative lack of insight into the perceptual qualities of reverberation placement, we decided that a small-scale pilot study would be the most appropriate initial experiment. While our experimental sample is not representative of our target demographic as a whole, based on our experimental focus, we do not anticipate significantly different results.

#### 3.1.1. Sample Population

We used a subject pool consisting of 25 members of the Carnegie Mellon University undergraduate population. This convenient sample allowed us to quickly gather data while maintaining a well-defined reference population. The final sample demographics reflect the Carnegie Mellon undergraduate community, with an approximately 60% male and 25% minority makeup. All participants were between the ages of 18 and 23. Subjects were not screened based on other demographics such as musical background.

#### 3.1.2. Sound Samples

For our test, we generated three sound samples for our subjects to compare. All three were based on the same samples of a 30-second jazz excerpt consisting of drum set, contrabass, and saxophone, all recorded with close microphones to minimize cross-source contamination. The samples were chosen because we felt that a non-classical source would result in a more pronounced sonic differentiation between instruments, while the jazz idiom

also requires a “live” enough feel that reverberation-based placement in a hall would be an appropriate effect.

To create our samples, we wrote a Nyquist-based [3] FFT convolution algorithm, which was then used to convolve hall-measured impulse response data with the dry jazz samples. These samples were then used to create three variations. The first, called *mono*, is a single-channel sample in which all three instruments are convolved with hall-center impulse responses. The second, referred to as *panned*, is a stereo sample in which all three instruments are convolved with the hall-center impulse response, and then panned such that the drums are center, the bass 80% right, and the saxophone 80% left. The final sample, called *placed*, convolves each instrument signal with a different impulse response: a center-based impulse response with the drum set, an audience-perspective right impulse response with the bass signal, and an audience perspective left impulse response with the saxophone signal.

At the highest granularity, the resulting sound samples are all reverberation-wet jazz performances, identical except for techniques regarding instrument placement in the stereo field. The samples were also normalized to peak at 0 dB so as to have matching volume levels. Upon initial listening by the investigators, the reverberation-placed sample seemed to display a richness lacking in the other two samples. The pilot study would later corroborate this subjective observation.

The impulse responses themselves were recorded via a microphone array located in the audience at the center of the concert hall. The venue chosen was the 200 seat Recital Hall located at the School of Music, University of Victoria, Canada. The responses were measured using a swept sine wave through a microphone array and repeated at three locations on the stage [8]. This resulted in an array of 7 different impulse responses for each location on the stage. For our simple stereophonic setup for this experiment, we chose simply the left and right impulse responses (2 of the 7 measured responses) for each of the 3 locations, corresponding to stage right, stage left and stage center. Other measured stereo impulse responses are available for a variety of concert halls and other venues [2]; however, these measurements typically do not include multiple locations on stage, and thus cannot be used for the *placed* variation in this experiment.

#### 3.1.3. Questionnaire

To compare the sound samples objectively, we developed a battery of comparative questions to grade the sound samples. The three categories of comparison were “realism,” as defined by the sample’s likeness to a live performance, “sound quality,” and simple personal preference. The format of the questionnaire was to ask the listener to listen to two sound samples consecutively, and then compare them on the three selected attributes. Each

sample was paired with every other sample, making for a total of three individual listening tests. To reduce bias, the order of the sample pairings was randomized as well as the play order within a given sample pair.

Due to concerns about the ability of all subjects to distinguish between the samples, the realism and quality questions asked for a simple pair-wise comparison to indicate which of the two samples the subject preferred across the realism, quality, and overall preference metrics described above. The preference question also asked for a comparison, but also allowed for answers of “I have no preference” and “I could not tell a difference.” In retrospect, listeners did not appear to have great difficulty in distinguishing the samples, with less than 6% of respondents selecting “no preference” or “no difference.”

### 3.2 Experimental Administration

The experiment was administered over the course of a weekend to all 25 subjects. Administration of the study was not difficult due to the brevity and subject matter of the experiment. The study proceeded in a randomized single-blind fashion, on one of two reference systems<sup>2</sup>. Regarding volume, listeners were asked to initially adjust the volume to preference, and then attempt to remain consistent throughout.

#### 3.2.1. Process

The study involved, first, a principal investigator providing the consent form and explaining that the study intended to compare several reverberation techniques, and that the listeners would be asked to listen to several jazz excerpts, identical except for the reverberation applied. The participants were then allowed to look over the questionnaire, but the investigator provided no interpretation as to the meaning of each question or questions regarding sample specifics.

At this point, the investigator played the first sample, identified only by a number, then the second sample. After this, the subject would record their results on the questionnaire, but the sound samples would not be replayed. The process was then repeated for the other two pairs of sound samples, the end result being that each subject would listen to each sound example twice and compare each to the others. After collecting the questionnaire, the investigators provided a brief explanation of the actual experimental intent and identified the sound samples by technique applied.

<sup>2</sup> Both systems were laptop PCs, one with Sony MDR-V500 headphones, and the other with Koss UR-40 headphones

#### 3.2.2 Data Analysis

For a study of this size, bias due to random variation in samples is a real concern. As such, we feel that it is important to include confidence intervals along with our proportion averages so as to accurately reflect the variability of our pilot study. For this study, we considered the experimental results to be drawn from a binomial distribution, and we calculated confidence intervals based on a normal approximation of this distribution [1]. The binomial distribution assumes that each experimental trial has only two outcomes; to match this model, the preference calculations dropped “no preference” and “no difference” responses.

For example, of the 25 participants, 8 perceived *panned* as sounding more realistic than *mono*. To compute the  $\alpha=.95$  confidence interval for realism, *panned* vs. *mono*, we simply used the binomial confidence interval formula for proportions:

$$CI = p \pm 1.96\sqrt{(p(1-p)/N)} \quad (1)$$

Here  $p = (8/25) = .32$  and  $N = 25$ . Thus,

$$CI = .32 \pm 1.96\sqrt{((.32(1-.32)/25)} \quad (2)$$

$$CI = .32 \pm .182 = [.137, .503] \quad (3)$$

Now we can interpret these data by saying that with 95% confidence, the true population proportion preferring *panned* to *mono* falls between 0.135 and 0.503, taking our sample size into account.

## 4. EXPERIMENTAL RESULTS

Our experimental results seem to point in favor of location-based reverberation for instrument placement based on the metrics of both sound quality and personal preference. Realism does not result in as conclusive of a result, but the data yields valuable insights.

	Panned vs. Mono	Placed vs. Mono	Placed vs. Panned
Realism	p = .32 [.137, .503]	p = .52 [.324, .716]	p = .68 [.497, .863]
Quality	p = .72 [.497, .863]	p = .84 [.696, .984]	p = .64 [.452, .828]
Preference	p = .57 [.363, .768]	p = .70 [.508, .884]	p = .68 [.497, .863]

**Table 1.** Aggregated means and confidence intervals for proportion preferring the first listed sound clip in each cell.

### 4.1. Realism

In this study, we defined realism as “likeness to an actual live performance.” Interestingly, there does not appear to be a strong consensus on what a live performance sounds like. Each pair-wise comparison of realism resulted in a confidence interval that included .5, the null hypothesis that there is no perceived realism difference between the

samples (see Table 1). Nevertheless, .68 rated the *mono* sample as more realistic than *panned*, and .68 rated the *placed* sample as more realistic than *panned*. This may be a reflection of a lack of realism in the *panned* sample, where the stereo spread could have been too wide to be considered realistic. Conversely, it may simply reflect a tendency of the sample population to feel that smaller stereo spreads best reflect the experience of a live performance, especially over headphones, which can exaggerate panning effects.

The other interesting observation about realism is the fact that the proportion preferring *placed* to *mono* was .52, almost exactly the null hypothesis. While the other two pairs were barely out of the 95% confidence range, it appears that our sample population could not distinguish between the two with regards to realism. We hypothesize that this indicates that the stereo spread effect is potentially a major determining factor in causing listeners to perceive a recording as realistic.

#### 4.2. Sound Quality

As opposed to realism, our investigation found much stronger support for location-based reverberation placement with regards to “sound quality.” Here, *mono* fared the worst, with .72 of the population preferring *panned*, and an extremely high .84 of the population preferring *placed*. In fact, despite the small sample size, the *placed* versus *mono* confidence interval, [.696,.984], is highly significant, and the *placed* versus *panned* interval, [.452,.828], which only barely contains the .5 null hypothesis, is close enough to significant to motivate a larger study to determine if location-based reverberation is truly a higher-quality placement technique than panning.

One other interesting trend to note is the relationship between realism and quality for each of the three pairs. The observed relationships vary in counter-intuitive ways. Quality and realism correlate positively for *placed* versus *panned*, while they correlate negatively for *panned* versus *mono*. Finally, subjects decisively find *placed* to be of higher quality than *mono*, but seem to be unable to decide which is more realistic. With our sample size, it is entirely possible that these trends are just random noise, but their further exploration on a larger sample could prove instructive.

#### 4.3. Personal Preference

The final metric is overall personal preference of the various sound samples. This measure shows the greatest advantage for convolution reverb placement. Subjects preferred *placed*, with .70 rating it over *mono* and .68 rating it over *panned*. Even with only 25 participants, the *mono* comparison is significant at the  $\alpha=.95$  level, and the *panned* comparison just barely misses this level of

significance (see Table 1). We feel such a consistent result in favor of convolution placement is solid evidence that the technique is a viable improvement over current post-processing effects. More subjects and a larger variety of sample material would likely serve to add weight to this judgement.

In addition to these results, we find it interesting that preference seemed much more split when comparing *mono* and *panned*. Subjects preferred *panned*, but only .57 rated it over *mono*. If it really were true that the increased perception of realism in *mono* somehow cancelled out the increased sound quality with *panned*, this would prove to be another advantage for location-based reverberation placement, which seems to be able to combine the best qualities of both other methods. That said, this interpretation seems unlikely, and a much larger pool of subjects and samples would be necessary to give it much credence. The strongest indication of this pilot study is the overall preference for location-based placement over other techniques.

### 5. CONCLUSION

Judging by this pilot study, the potential impact of location-based reverberation techniques on the recording industry is large. If this technique is indeed perceived to be of better quality and more preferable than current recording techniques, then there is clear potential for commercial viability. The major logistical obstacle to overcome would be to gather a much larger pool of impulse response data for use by industry standard convolution reverberation plug-ins such as the Waves IR1. Since plug-ins of this sort already rely on hall-measured impulse-response data, the burden of measuring a larger number of instrument/listener location pairs should not be too prohibitive. Location-based reverberation has the potential to lend itself as a relatively inexpensive and effective post-processing technique that can be used in today’s stereophonic applications to greatly enhance the psycho-acoustical experience for the listener.

The results of our single-blind pilot study clearly warrant further investigation. Within the bounds of our sample size and limited demographic, our results point in favor of location-based reverberation placement. The average listener’s preference to the location-based reverberation technique demonstrates the technique’s potential viability in the commercial realm. It is therefore advisable that studies regarding this technique should be continued in larger and more controlled environments across a wider range of sound samples. We expect that larger studies will generate conclusively positive results and that location-based reverberation placement has the potential to become an industry standard technique for artificial reverberation and localization in stereophonic recordings.

## 6. REFERENCES

- [1] Agresti, Alan. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, New York, 1996.
- [2] Audio Ease Impulse Responses. (Online at: <http://www.audioease.com/IR/index.html>)
- [3] Dannenberg, R. "Machine Tongues XIX: Nyquist, a Language for Composition and Sound Synthesis," *Comp. Music Journal*, 21(3), Fall 1997, pp. 50-60.
- [4] Griesinger, D, "Beyond MLS occupied hall measurement with FFT techniques," *101st Audio Eng. Society Convention*, Preprint 4403, Oct. 1996.
- [5] Pulkki, V, *Spatial sound generation and perception by amplitude panning techniques*, Ph.D. Dissertation, Helsinki Univ of Technology, 2001.
- [6] Reller, C.P.A, Jawksford, M.O.J., "Perceptually motivated processing for spatial audio microphone arrays," *115th Audio Engineering Society Convention*, preprint 5933, October 2003.
- [7] Youngblut, C., Johnston, R., Nash, S., Wienclaw, R., and Will, C. *Review of Virtual Environment Interface Technology*. IDA Paper P-3186. Alexandria, VA: Inst. for Defense Analysis (IDA), Mar. 1996. (Online at: <http://www.hitl.washington.edu/scivw/scivw-ftp/publications/IDA-pdf/>)
- [8] Li, Y., Driessen, P.F., Tzanetakis, G., Bellamy, S. "Spatial Sound Rendering Using Measured Room Impulse Responses," *Signal Processing and Information Technology, 2006 IEEE International Symposium on ISSPIT 2006*, Aug. 2006, pp. 432-7.